- Search History will be lost after eight hours of inactivity.
- To combine searches use # before search number, e.g., #2 AND #6.
- Search numbers may not be continuous; all searches are represented.
- Click on query # to add to strategy

Search	Most Recent Queries	Time	Result
#12	Search **cantor**	10:58:09	246
#10	Search **cantor and haplotype**	10:58:05	0
#11	Search **ding and haplotype**	10:57:59	40
#9	Search **jeffreys and wilson and neumann**	10:56:38	3
#8	Search **ross and hall and haff**	10:55:53	3
#7	Search **rein and zorbas** Field: **All Fields**	10:55:17	8
#5	Search **rein and methylation**	08:12:32	13
#3	Search **epigenetic haplotypes methylation**	08:10:47	6
#2	Search **epigenetic haplotypes**	08:10:35	22

EAST Search History

Ref #	Hits	Search Query	DBs	Default Operator	Plurals	Time Stamp
S26	0	ruaono.in.	US-PGPUB; USPAT; USOCR; EPO; JPO; DERWENT	OR	ON	2006/03/15 07:53
S27	123	ruano.in.	US-PGPUB; USPAT; USOCR; EPO; JPO; DERWENT	OR	ON	2006/03/15 07:53
S28	0	ruanl2 and haplotype	US-PGPUB; USPAT; USOCR; EPO; JPO; DERWENT	OR	ON	2006/03/15 07:53
S29	4	S27 and haplotype	US-PGPUB; USPAT; USOCR; EPO; JPO; DERWENT	OR	ON	2006/03/15 07:55
S30	2	"6931326"	US-PGPUB; USPAT; USOCR; EPO; JPO; DERWENT	OR	ON	2006/03/15 09:27
S31	1	2001-071535.NRAN.	DERWENT	OR	ON	2006/03/15 08:57
S32	2	single adj molecule adj dilution	DERWENT	OR	ON	2006/03/15 08:57
S33	2	"20020081598"	US-PGPUB; USPAT; USOCR; EPO; JPO; DERWENT	OR	ON	2006/03/15 13:22
S34	10	ding.in. and cantor.in.	US-PGPUB; USPAT; USOCR; EPO; JPO; DERWENT	OR	ON	2006/03/15 13:37
S35	123	ruano.in.	US-PGPUB; USPAT; USOCR; EPO; JPO; DERWENT	OR	ON	2006/03/15 13:24
S36	0	S35 and multiplex	US-PGPUB; USPAT; USOCR; EPO; JPO; DERWENT	OR	ON	2006/03/15 13:24
S37	0	S35 and hplotype	US-PGPUB; USPAT; USOCR; EPO; JPO; DERWENT	OR	ON	2006/03/15 13:24
S38	4	S35 and haplotype	US-PGPUB; USPAT; USOCR; EPO; JPO; DERWENT	OR	ON	2006/03/15 13:24

EAST Search History

S39	5	S35 and (haplotype or genotype)	US-PGPUB; USPAT; USOCR; EPO; JPO; DERWENT	OR	ON	2006/03/15 13:28
S40	46	jeffreys.in. and neumann.in.	US-PGPUB; USPAT; USOCR; EPO; JPO; DERWENT	OR	ON	2006/03/15 13:37
S41	8	jeffreys.in. and neumann.in. and sequence	US-PGPUB; USPAT; USOCR; EPO; JPO; DERWENT	OR	ON	2006/03/15 13:38

```
d his full

     (FILE 'HOME' ENTERED AT 11:01:10 ON 16 MAR 2006)

     FILE 'MEDLINE, AGRICOLA, CABA, CAPLUS, BIOSIS, DISSABS, EMBASE,
     1MOBILITY' ENTERED AT 11:01:36 ON 16 MAR 2006
                E CANTOR C R/AU 21
L1             1 SEA ABB=ON  PLU=ON  "CANTOR CAHRLES R"/AU
                D L1 1- TI
L2            19 SEA ABB=ON  PLU=ON  ("CANTOR C R"/AU OR "CANTOR C R C R"/AU OR
                "CANTOR CAHRLES R"/AU OR "CANTOR CH R"/AU OR "CANTOR CHARLES"/A
                U OR "CANTOR CHARLES R"/AU OR "CANTOR CHARLES ROBERT"/AU) AND
                HAPLOTYPE
L3             9 DUP REM L2 (10 DUPLICATES REMOVED)
                D L3 1- TI
                D L3 1- IBIB ABS
                E DING CHUNMING/AU 21
L4             9 SEA ABB=ON  PLU=ON  ("DING CHUNMIN"/AU OR "DING CHUNMING"/AU)
                AND HAPLOTYPE
L5             5 DUP REM L4 (4 DUPLICATES REMOVED)
                D L5 1- IBIB ABS
L6         83715 SEA ABB=ON  PLU=ON  (HAPLOTYPE OR GENOTYPE) AND (SINGLE OR
                DILUTION OR DILUTED)
L7           264 SEA ABB=ON  PLU=ON  L6 AND (SINGLE MOLECULE) OR (SINGLE
                TEMPLATE)
L8           136 DUP REM L7 (128 DUPLICATES REMOVED)
L9            80 SEA ABB=ON  PLU=ON  L8 AND PY<2003
                D L9 1-30 TI
                D L9 5 IBIB ABS
L10           12 SEA ABB=ON  PLU=ON  HAPLOTYPE AND TEMPLATE AND (DILUT?)
                D L10 1- TI
L11            5 DUP REM L10 (7 DUPLICATES REMOVED)
                D L11 1- IBIB ABS
L12           13 SEA ABB=ON  PLU=ON  (HAPLTOYPE OR GENOTYPE) AND ((SINGLE
                MOLEUCLE) OR (SINGLE TARGET))
                D L12 1- TI
```

ordered C. Ding dissertation (Boston U) from STIC-ILL on 3-16-06

Genomics & Proteomics

Helping SNPs to Speak Up Louder Than Before

Studying genetic variation generally shows much promise, but finding the phase or haplotype of a set of SNPs is tricky business. Experimental approaches are being explored that could facilitate haplotype analysis for SNPs separated by longer stretches

By Vivien Marx

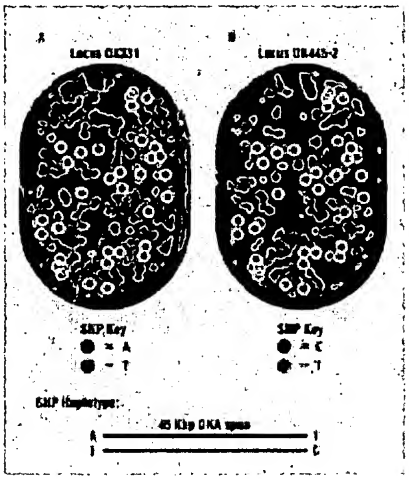
If single-nucleotide polymorphisms (SNPs) offer quiet evidence of genetic diversity, then haplotypes crank up the volume. Researchers are working on better methods to let them hear the precise tune that haplotype blocks of genomic variation play. Single-molecule techniques, for example, permit a new opportunity for long-range haplotyping. Some techniques being explored involve the co-amplification of two loci on the same molecule, i.e., two 100-base-pair loci rather than one 20-kilobase (kb) stretch that has both SNPs on it.

In a given cell, haplotyping, in particular resolving the phases of several markers, is a challenge. If a region has three heterozygous markers, that means eight possible haplotypes. Pedigree information and knowledge of computational mechanisms can help hone in on those possibilities. Other methods that do not require pedigree information involve haplotyping the direct molecular way, by separating the homologous DNAs before genotyping. Then comes DNA cloning, hybrid construction, and single-cell long-range PCR, all of which are time-consuming and limited to short genomic stretches. Computational methods, such as inference from unphased data, have limitations, for example, when it comes to genomic length.

It all depends on the locus, but computational inference is a hard problem, says Robi Mitra, PhD, assistant professor of Genetics at Washington University School of Medicine, St. Louis. "In some haplotypes," he says, "just by eye you see where recombination events occur, that's a no-brainer." And then there are areas where heterogeneity and the number of haplotypes, the recombination, all make the task computationally hard with plenty of probabilistic solutions given the data.

The problem with statistical and computational methods, says Charles Cantor, PhD, chief scientific officer of Sequenom, Inc., San Diego, is their mistakes. "If you are going to do haplotyping as part of a population analysis," he says, "you may be tolerant of a certain percentage of mistakes, but if you are going to haplotype in a clinical diagnostic setting, you are not tolerant to any those kinds of things."

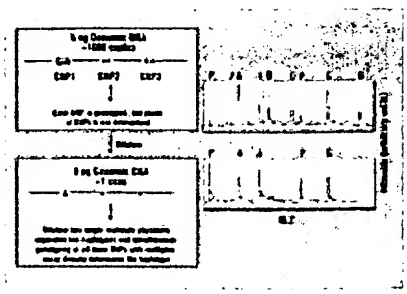
An approach currently being explored in a number of labs is haplotyping with single-molecule amplification. One method, molecular colony technique, developed in Alexander Chetverin's lab at the Institute of Protein Research of the Russian Academy of Sciences involves amplification in acrylamide. There is BEAM-ing, a word derived from the method's components of beads, emulsion, amplification, and magnetics. Developed by Devin Dressman, PhD, and colleagues in Bert Vogelstein's lab at Johns Hopkins, it involves converting each DNA molecule into a single magnetic particle to which thousands of identical copies of DNA are bound.



click the image to enlarge

Phasing of SNPs separated by a genomic distance of 45 kb. The scientists amplified and genotyped SNPs DK331 and DK445-2 by single-base extension (SBE). Some 153 polonies were genotyped at the DK331 locus (A) and 175 polonies were genotyped at the DK445-2 locus (B). Some 34 overlapping polonies were identified (white circles indicate polonies), and 32 of them show that the A variant at SNP DK331 is in cis with the T variant of SNP DK445-2 (LOD=20.3). [Source: Robi Mitra, reprinted with permission from *Proc. Natl. Acad. Sci. U.S.A*, vol. 100, p. 5920 (2003)]

Polony: a sausage and a colony



click the image to enlarge

Upper right: Traditional genotyping methods use 5 ng of genomic DNA, or about 1,600 copies of genomic templates. They yield genotypes of each SNP but do not determine SNP phases. Lower right: Simultaneous genotyping of several markers using multiplex assays with single-copy DNA molecules makes haplotyping analysis possible. The two alleles can be physically separated at extremely dilute DNA concentrations. This approach does not amplify the entire haplotype block, only around 100 base pairs near each SNP are amplified. P peaks are from unextended primers. [Source: Charles Cantor, Chunming Ding, reprinted with permission from *Proc. Natl. Acad. Sci. U.S.A.*, vol. 100, p. 7450 (2003)]

Another method is called polymerase colony amplification or "polony" technology for short. Mitra started working on polonies as a graduate student in George Church's lab at the Lipper Center for Computational Genetics in the Department of Genetics at Harvard Medical School. It was Church's idea to amplify single molecules in acrylamide and they subsequently worked out different applications such as sequencing and haplotyping.

"This is a way to amplify a large number of single DNA molecules," says Mitra. Much like the apples and the proverbial tree, due to the constraints of the acrylamide, each of the amplification products remains localized near its parent molecule. Every molecule in the colony has originated from the same parent molecule, so large numbers of PCR reactions can happen at the same time while keeping the resulting DNA separate.

Although these techniques represent the beginning stages, interviewed scientists point out they all show the power of analyzing single molecules rather than populations of molecules.

To illustrate polony haplotyping, Mitra, Church and colleagues performed an in-gel PCR reaction with a small amount of patient DNA with two pairs of PCR primers capable of amplifying one of the SNPs. Thus, two loci from a single DNA molecule are amplified. The gel's structure keeps the amplification products close by the chromosome as two overlapping polonies. One strand of the amplified DNA is attached to the matrix via a covalent bond permitting removal of the other one. The polonies were genotyped and the phase of the SNPs determined by the overlapping polonies. Locus-specific primers carrying fluorescent dyes helped to visualize them.

The team managed to determine the genotype and phase of three different pairs of SNPs and, in one case, the SNPs were at a distance of 45 kb. The distance, says Mitra, is because they used DNA prepared by standard methods, which tends to be sheared at about 50 kb sizes. The distance limit is due to their DNA source, he says. "Basically," Mitra says, "there is no reason that with this technique we can't go longer with just a different purification method."

Another shot at the long-range

Cantor, along with Chunming Ding, PhD, of the Bioinformatics Program at the Center for Advanced Biotechnology, Boston University, proposes M1-PCR as their road to more success for direct molecular haplotyping of long-range genomic DNA.

As Ding explains, very little DNA is needed, "only 3 pg, which is 1,000 times less than for a typical genotyping assay." No DNA from parents or siblings is required and the assay works on DNAs extracted from peripheral blood, thus the time consuming steps of cell line construction or DNA cloning are avoided.

In M1-PCR, the "M" stands for multiplexing and the "1" for single-copy DNA molecules. The process entails taking a DNA sample and diluting it to single-copy, which separates the two homologous DNAs. Then comes the direct multiplex genotyping of several markers with Sequenom's MALDI-TOF MS-based MassArray system. A built-in software tool automatically categorizes each analyzed sample as to whether one or both alleles are present.

Applying this system, the scientists were able to haplotype several polymorphic markers separated by as many as 24 kb. "We can multiplex and so when we start with a single molecule, we do a multiplex of short PCRs flanking the particular SNP," says Cantor. "We amplify just around the SNP instead of doing the long stretch." Only approximately 100 base pairs of the region around each SNP are amplified through PCR in this process. By integrating M1-PCR into the MALDI-TOF system, the scientists say that high-throughput direct molecular haplotyping of a few thousand assays a day can be obtained.

As part of their experiments, Ding and Cantor compared the haplotypes determined by their method with those determined through genotyping and with pedigree information. At first, their method yielded a number of

incomplete haplotype calls, which stood out. "When there are breaks in the molecule, what we see is an incomplete genotype or haplotype, Cantor says. "And then we know that our single molecule was missing an end."

The trick, says Cantor, is to have enough replicates. In order to counter this effect, they used more replicates, 10 to 14 replicates, of single-molecule PCR for each individual. "This is not a problem, since most of the cost of doing this is in the sample preparation," he says. "Once you are diluting the sample to one molecule, the cost is insignificant." Distance between SNPs is not a limiting factor, the researchers say, as long as not many copies of the genomic DNA have breaks between the markers.

Cantor and Ding say that their technique is applicable to clinical settings where speed, sample size, and ease of use are all key. They realize it must be validated on lots of samples. "We have to make sure that an average lab can do it, not just a lab with our experience," Cantor says.

There was a prophet

Back in the 1980s, Gualberto Ruaño, MD, PhD, former CEO of Genaissance Pharmaceuticals, who founded Genomas LLC, both of New Haven, Conn., explored the single-molecule dilution, but found he could not get two or three amplifications at the same time. "We always knew the limit of this method would be imposed by the structure of the DNA molecule," he says.

In his 1990 paper, he outlined how haplotypes could be applied, much of which has come to pass. "Iam a good prophet," Ruaño says. "It is extremely satisfying to see the technique advancing." Part of that push was to found Genaissance Pharmaceuticals Inc., to integrate gene variation into the development of drugs. He does regret that he was "too naïve" to patent his methods. "Then again, the patents would be close to expiring now," he says.

Into the clinic, slowly

Collecting families for analysis is time-consuming, Cantor points out, and may be unrealistic for a clinical diagnostic setting in which rapid decisions are needed. Chunming Ding offers a possible scenario: "For example, the R117H mutation in the cystic fibrosis transmembrane receptor (CFTR) gene shows mild effect with the 5T mutation, and severe effect when the 5T mutation is present on the same chromosome. Thus, a haplotype of R117H-5T is important for clinical applications to determine the severity of the prognosis of this type of cystic fibrosis." There are other diseases in which a second mutation on the same chromosome can change the disease manifestation from the first mutation, Ding says.

Mitra explains that the method he explores as well as the one applied by Cantor and Ding share the novel idea of co-amplification of two loci on the same molecule. They both carry the message: "Don't be limited by your amplicon, don't try to PCR amplify both SNPs together, just co-amplify your SNPs in such a way that if they are from the same DNA molecule they are going to mingle together . . ." Mitra says. In his view, the studies show "the only limit you are putting on your distance now is 'can you keep your DNA intact?' "

Diagnosing the future

According to Cantor, his company believes that "mass spec has specific diagnostic advantages for nucleic acids over methods that existed before." Single-molecule haplotypting is one of them, so the company is wondering if the advantages are sufficient for "an entry into the diagnostic industry."

Glimpsing the Future ...

SNPs are not disease indicators per se. When specific SNPs are present on the same chromosome, disease manifestation is not the same as when the SNPs are on different chromosomes. If haplotyping is to make its way to the clinic, robust, scalable techniques are needed. According to scientists interviewed for this story, single-molecule analysis may be a promising approach, particularly co-amplifying two or more loci on one molecule.

For clinical applications, molecular haplotyping offers the potential of accuracy and low cost. For how many clinical diseases do you need to know the haplotype? "Right now, there are not that many," Mitra says. The belief is, though, that number will start to explode as mining of the human genome proceeds.

Organizations mentioned in this article:

Charles Cantor, PhD
Chunming Ding, PhD, Boston University Center for Advanced Biotechnology
Chunming Ding
Sequenom Inc.
Sequenom high-performance genotyping
Robi Mitra, PhD, Washington University School of Medicine

Polony protocols
Polony discussion forum
Polony technology
Devin Dressman, PhD; Bert Vogelstein, PhD
Genomas LLC, Gualberto Ruaño, PhD
Alexander Chetverin, PhD, Laboratory of Viral RNA Biochemistry, Russian Academy of Sciences